PEER REVIEW HISTORY

ARTICLE DETAILS

TITLE (PROVISIONAL)	Implementation, efficacy, and cost-effectiveness of the Unified Protocol in a Blended Format for the Transdiagnostic Treatment of Emotional Disorders: study protocol for a multicentre, randomized, superiority controlled trial in the Spanish National Health System
AUTHORS	Osma, J; Martínez-García, Laura; Peris-Baquero, Óscar; Navarro-Haro, María Vicenta; González-Pérez, Alberto; Suso-Ribera, Carlos

VERSION 1 – REVIEW

REVIEWER	Estupiñá-Puig, Francisco Complutense University of Madrid, Dpt. of Personality, Evaluation & Clinical Psychology
REVIEW RETURNED	26-Jul-2021

GENERAL COMMENTS	The study is pertinent and clearly relevant due to the demand of mental health received at the Spanish NHS and it's limited capacity to offer therapy. In fact, other trials are currently assessing the effectiveness of transdiagnostic, evidence based treatments at the Spanish NHS, as the PsicAP study, (cfr. Cano-Vindel et al., 2016) while this is the first that assesses this approach in a blended fashion and within individual therapy, the format preferred by the population. The protocol is well defined and generally solid. The research group behind the study is well stablished and currently spearheads the dissemination of the Unified Protocol in Spain, with some other RCTs currently publishing results on it's behalf (Osma et al., 2021). Nonetheless, some minor concerns might be raised about some traits of the study that might become problematic. I would like to address these issues for the authors, in case they find them useful. First, I have some concerns about the sample size. The prevision of a 15% drop out considered by the authors seems too optimistic. In fact, in previous works from the same group (Osma, Castellano, Crespo, & García-Palacios, 2015) abandon rates at post treatment have been around 25%, which is more in line with the usual attrition found in effectiveness studies, with a greater drop out rate as follow up measures go by. Also, no rationale or evidence base is provided for the effect size of d = .30 that is mentioned as the basis for the sample size requirements; while this is considered a conservative estimate by the authors, I worry that the TAU provided, if sessions are offered weekly or biweekly (as they should), should offer close results. Also, while the authors offer a calculation on the sample required, they offer no information on the number of candidates to be screened to achieve the desired sample. Given that the exclusion criteria include suicide risk and substance use, and also, having

	received therapy in the previous five years (which might be as high as 40% in clinical samples, cfr. Estupiñá, 2016), I'm afraid that a significant amount of candidates will be screened out, since these criteria, as defined in the study, will be met by a significant number of them. Also, no information is provided on how is suicide risk or substance use going to be screened. Substance abuse might also provide a more sensitive exclusion criteria. While it is true that the group has provided large samples in previous studies (Osma et al., 2021), I think that a more accurate estimate of the sample size to be screened could help the study to achieve it's objectives. Second, regarding the data collection, the use of the UP-APP to collect measures from the UP group guarantees a good ratio of completion within timing, provided a reliable app is developed; while I'm sure the authors have already considered this, a pilot study on the app is highly recommended. Also, it might have research interests by itself; we're currently lacking studies on the advantages of employing apps to collect data compared to traditional assessment. On the contrary, the TAU group might fare not so well; measures are planned to be collected weekly, but session frequency is left open to the "characteristics of their centres" (p. 11). Moreover, key information about the development of the TAU therapy is planned to be acquired at the end of the sessions. This poses some threat to the appearance of missing data and memory biases. I'm concerned that these issues might introduce problems in the comparison of both treatments, which might prevent solid conclusions being drawn on the role of the UP plus blended methodology due to disparities in the data available for both treatments, or the amount of therapy received. Also, I'm sure that the authors realize that the comparison between UP blended therapy and TAU makes it difficult to conclude the separate role of the UP protocol or the blended element in an eventual superiority of the UP plus app group. Third, the last observation carried forward (LOCF) strategy to approach missing data mentioned in p. 12 has been criticized as lacking a sound statistical basis (Leon et al., 2006; Sanz & García-Vera, 2013). I suggest using only modelling techniques for missing data imputation whenever possible, given an acceptable percentage of missing data (less than 20-30%) and a random pattern of missing data. Finally, I consider that the calculation of a Reliable Change Index (RCI) and a Reliable Recovery Index (RRI), as proposed by Jacobson and Truax (1991), would add a powerful yet simple tool to assess – and communicate – the effectiveness of both interventions, beyond the already powerful Cohen's d effect size. Even with the employment of the MINI as recovery criteria, I encourage the authors to consider these two procedures for their data analysis strategy. References: Cano-Vindel, A., Munoz-Navarro, R., Wood, C. M., Limonero, J. T., Medrano, L. A., Ruiz- Rodriguez, P., . . . Santolaya, F. (2016). Transdiagnostic Cognitive Behavioral Therapy Versus Treatment as Usual in Adult Patients With Emotional Disorders in the Primary Care Setting (PsicAP Study): Protocol for a Randomized Controlled Trial. JMIR Res Protoc, 5(4), e246. doi:10.2196/resprot.6351 Estupiñá Puig, F. J. (2016). Práctica clínica basada en la evidencia para el tratamiento psicológico de los trastornos depresivos: utilidad clínica y coste-efectividad. Universidad Complutense de Madrid, Madrid. Retrieved from http://eprints.sim.ucm.es/37648/

	Jacobson, N. S., & Truax, P. (1991). Clinical significance: A statistical approach to defining meaningful change in psychotherapy research. Journal of Consulting and Clinical Psychology, 59(1), 12-19. Retrieved from http://psycnet.apa.org/journals/ccp/59/1/12/ Leon, A. C., Mallinckrodt, C. H., Chuang-Stein, C., Archibald, D. G., Archer, G. E., & Chartier, K. (2006). Attrition in randomized controlled clinical trials: methodological issues in psychopharmacology. Biological Psychiatry, 59(11), 1001-1005. doi:10.1016/j.biopsych.2005.10.020 Osma, J., Castellano, C., Crespo, E., & García-Palacios, A. (2015). The unified protocol for transdiagnostic treatment of emotional disorders in group format in a spanish public mental health setting. Behavioral Psychology 23(3), 447-466. Osma, J., Peris-Baquero, J. Suso-Ribera, C., Farchione, T. & Barlow, D. (2021) Effectiveness of the Unified Protocol for transdiagnostic treatment of emotional disorders in group format in Spain: Results from a randomized controlled trial with 6-months follow-up. Psychotherapy Research. Advance online pubication. DOI: 10.1080/10503307.2021.1939190 Sanz, J., & García-Vera, M. P. (2013). Rendimiento diagnóstico y estructura factorial del Inventario para la Depresión de Beck–Segunda Edición (BDI-II) en pacientes españoles con trastornos psicológicos. Anales de Psicología, 29(1), 66-75. doi:10.6018/analesps.29.1.130532

REVIEWER	Klein, Corinna University of California Santa Barbara
REVIEW RETURNED	20-Aug-2021

GENERAL COMMENTS	Overview I appreciate the opportunity to review the protocol: "Implementation, efficacy, and cost-effectiveness of the Unified Protocol in a Blended Format for the Transdiagnostic Treatment of Emotional Disorders: study protocol for a multicentre, randomized, superiority controlled trial in the Spanish National Health System." This protocol presents a multicenter, randomized, superiority, clinical trial evaluating the UP transdiagnostic treatment by comparing CBT (which they define as treatment as usual) with a blended format UP (delivered through both in-person and online formats). I believe that this study could make significant contributions to the literature and our understanding of how blending in-person clinical services with app-based treatment may support patients and providers while alleviating burden for mental health systems. However, several clarifications and revisions should be made to the protocol, delineated below. I appreciate that the researchers intend for the study to be as naturalistic as possible, in order to ensure accurate representation of services delivered in public settings, however, as the authors indicate, this adds many confounding variables that may obfuscate study results. Additionally, there seem to be other more significant confounds that make it difficult to disentangle whether differences between study condition are attributable to differences in treatment (CBT vs UP) or to the blending of in-person and online services. Finally, blended treatment is proposed as a way of alleviating burden on care systems and saving time for clinicians, but in this study online interventions seem to be more of a supplement to treatment than a replacement for some in-person sessions. Additional comments follow:

	Abstract The final bullet point in strengths and limitations is syntactically incorrect. Introduction Pg 5 Line 47: Word “collapsed” should be replaced with “overwhelmed” or a synonym. Pg 6 Line 3: misplaced infinitive “it allows to simultaneously treat” or missing noun The introduction would benefit from more background information on blended treatments. It states that blended treatments are “dynamic and flexible because they allow using technology to motivate, monitor, give support, and treat patients,” but these are a broad array of benefits and outcomes that seem to be referenced together without citing research that clearly demonstrates these benefits. They should be parsed out more, since flexibility for clinicians is a different issue from monitoring and motivation. Please include information about other treatments that have been blended, and whether the blending usually includes supplemental support or the replacement of some in-person sessions with online modules. Pg. 6 line 40: “save time to the clinicians” should be reworded, and “w” should be replaced with a full word. Cost-efficiency is referenced as a study outcome in the introduction and title but not in the abstract. Pg 6. 50: “our goal will be tested” should be reworded for clarity (e.g. “our outcomes will be evaluated,” “our aims will be assessed,” “our study will be conducted”) Study Protocol I am struck again by the significant confound of providing two different treatment protocols, one structured and one unstructured, but suggesting that the superiority will be attributable to the blended format, or additional treatment, in one condition. Why not offer the UP to both groups so that any comparative benefits would more clearly be attributable to the additional online support? As the study is currently described, differences in condition could be due to the protocol (UP vs CBT), the presence or lack of structured protocol, the addition of an online element, or simply additional support. It would also be important to understand how therapists will be trained in the UP and how fidelity to the UP will be monitored to make sure the interventions (UP vs CBT) are, in fact, different. How will the researchers be ensuring that UP therapists are not providing UP to clients who are randomized to TAU? Line 20 “standard intervention” should be clarified as to which treatment group the authors are referring to. Sample Size I appreciate the description of how a sample size was generated. However, a dropout rate of 15% in a public mental health setting seems optimistic, given that reported dropout rates are typically double that. The authors should explain how they determined this

	dropout rate (is it specific to the centers where the study will be conducted, or to the Spanish healthcare system?). Fernández D, Vigo D, Sampson NA, Hwang I, Aguilar-Gaxiola S, Al-Hamzawi AO, Alonso J, Andrade LH, Bromet EJ, de Girolamo G, de Jonge P, Florescu S, Gureje O, Hinkov H, Hu C, Karam EG, Karam G, Kawakami N, Kiejna A, Kovess-Masfety V, Medina-Mora ME, Navarro-Mateu F, Ojagbemi A, O'Neill S, Piazza M, Posada-Villa J, Rapsey C, Williams DR, Xavier M, Ziv Y, Kessler RC, Haro JM. Patterns of care and dropout rates from outpatient mental healthcare in low-, middle- and high-income countries from the World Health Organization's World Mental Health Survey Initiative. Psychol Med. 2020 Apr 28:1-13. doi: 10.1017/S0033291720000884. Epub ahead of print. PMID: 32343221; PMCID: PMC8265313. Procedure How will the qualitative analysis of focus group data be collected? Under sample and recruitment, which mental health professionals will be responsible for assessing diagnoses? The referring ones or study therapists? Under randomization, the section about patients who refuse to participate in the study is confusing. Does this mean you will be collecting data on these patients as well? Regarding the stratified sampling strategy, how will the researchers account for individuals who fit into multiple diagnostic categories (i.e. GAD and depression) and are more severe in one or less severe in another? Please be more specific about “the recommended cut-off in the manuals.” Which measures are you using to determine severity of depression and/or anxiety, and what cut-offs will determine severity for stratified randomizing? How will researchers be simultaneously ensuring that 10 participants at each cite are assigned to each group (as is indicated in the flow chart). Pg. 11, line 16: typo: “this individual face-to-face appointments” (these). Pg. 11 line 19-20: typo in the following sentence: “clinicians will recommend the participants in the blended condition to work on…” p. 11 line 26: should read “prevents us from defining” In your description of the two treatments, pharmacological treatment is mentioned only under TAU. The inclusion criteria chart suggested that nobody would be excluded due to psychotropic treatment, but it is confusing that this is mentioned only under TAU. Presumably participants in either group may be receiving conjunctive pharmacological treatment. Please clarify. UP treatment group: The team has obviously not developed the app yet, and plans to design it in accordance with expressed preferences gathered in focus groups, but more information about what the app might involve would be useful. Will it have daily mood tracking or activities related to each UP module? Will it have UP homework assignments in digital format? For readers unfamiliar with the UP module, please provide examples of the types of engagement users will have with the app and more information about what the UP entails. Measures

	More information about all chosen measures would be beneficial to this protocol, including information on their validation, reliability, and general descriptions of the measure. For example, are the SUS and CEQ self-reports? Are they completed by therapists or administrators? What is the CSRI? Given the large number of outcomes being evaluated, the measures section would benefit from being subdivided into clearer sections (i.e. secondary outcomes could include Patient Outcomes, Implementation Outcomes, and Patient Satisfaction Outcomes). App Outcomes I believe that "without the need to ask the participant" intends to clarify that participants will not need to actively respond to questions about their use, but it reads as a lack of consent. I assume that participants will be informed of all data gathered by the app. Please clarify. Analyses Given the large number of possible confounds in this study (i.e. number of sessions, length of treatment sessions, frequency of appointments) and the large number of outcomes collected, what covariates will be included in your analyses? In the multi-level model, what will each level of the model be? Will models account for treatment center where interventions were received? Please provide further information on how relationship between cost of intervention and QALYs will be analyzed. Conclusions The study is presented in the conclusions as offering insight into a potential solution to long waiting lists, however this study does not include treatment time or other measures related to waitlists as primary study outcomes. This seems like a future direction rather than something that the study will "reveal." Appendices Pg. 23: Flowchart: This chart should have a title and heading. The table of assessments included an "intervention" assessment (t2) where the ODSIS and OASIS are administered. When will this occur? Or will these be provided throughout treatment, and if so how frequently? The study protocol suggests that these two measures will be delivered to APP users through the App. How will the TAU group receive these measures throughout treatment?

VERSION 1 – AUTHOR RESPONSE

Reviewer: 1

The study is pertinent and clearly relevant due to the demand of mental health received at the Spanish NHS and it's limited capacity to offer therapy. In fact, other trials are currently assessing the effectiveness of transdiagnostic, evidence based treatments at the Spanish NHS, as the PsicAP study, (cfr. Cano-Vindel et al., 2016) while this is the first that assesses this approach in a blended fashion and within individual therapy, the format preferred by the population.

The protocol is well defined and generally solid. The research group behind the study is well stablished and currently spearheads the dissemination of the Unified Protocol in Spain, with some

other RCTs currently publishing results on it's behalf (Osma et al., 2021). Nonetheless, some minor concerns might be raised about some traits of the study that might become problematic. I would like to address these issues for the authors, in case they find them useful.
First, I have some concerns about the sample size. The prevision of a 15% drop out considered by the authors seems too optimistic. In fact, in previous works from the same group (Osma, Castellano, Crespo, & García-Palacios, 2015) abandon rates at post treatment have been around 25%, which is more in line with the usual attrition found in effectiveness studies, with a greater drop out rate as follow up measures go by. Also, no rationale or evidence base is provided for the effect size of d = .30 that is mentioned as the basis for the sample size requirements; while this is considered a conservative estimate by the authors, I worry that the TAU provided, if sessions are offered weekly or biweekly (as they should), should offer close results.

Response: Thank you very much for raising this question. The rationale behind the relatively low dropout rates compared to the previous work from our group (despite this reference is from a non-controlled pilot study with 11 participants), as well as the proposed effect size comes from a literature review showing that blended interventions lead to significantly lower dropout rates (Erbe et al., 2017) and are superior to face-to-face treatments in several outcomes in patients with anxiety and adjustment disorder (Leterme et al., 2020).We now justify this with the mentioned references in the text.

Modified text in page 5 Lines 7-10: "The expected effect size and dropout rates come from studies showing that blended interventions lead to lower dropout rates [20] and better outcomes in patients with anxiety and adjustment disorder [18] when compared to face-to-face interventions."

Also, while the authors offer a calculation on the sample required, they offer no information on the number of candidates to be screened to achieve the desired sample. Given that the exclusion criteria include suicide risk and substance use, and also, having received therapy in the previous five years (which might be as high as 40% in clinical samples, cfr. Estupiñá, 2016), I'm afraid that a significant amount of candidates will be screened out, since these criteria, as defined in the study, will be met by a significant number of them. Also, no information is provided on how is suicide risk or substance use going to be screened. Substance abuse might also provide a more sensitive exclusion criteria.
While it is true that the group has provided large samples in previous studies (Osma et al., 2021), I think that a more accurate estimate of the sample size to be screened could help the study to achieve it's objectives.

Response: We want to thank the reviewer for noticing that the group has made an effort to provide large and representative samples in the past. This is also our intention in the present study, which is why sample size calculations were made. We understand that large samples are always preferable, but overestimating the sample size required for a study might also be detrimental in terms of clinician/researcher/participant burden, which is why we have relied on previous literature to calculate the sample size and provide an achievable, yet sufficient sample size to achieve our goals. In response to the number of participants who have to be evaluated and who may not meet the inclusion criteria, this number should be relatively low, considering that the research project will be carried out in specialized care units. In our previous project, out of 507 participants evaluated, only 19 (3.75%) did not meet the inclusion criteria. Taking these data into account and, as suggested by the reviewer, we have increased the number of candidates to be screened. We have modified this in the document and in the clinicaltrials registry record.
Modified text in page 5 Lines 3-10: "To calculate the required sample size, we used the G*Power software [21]. We obtained a sample size of 129 participants per condition with a 95% power, an alpha coefficient of 0.01, and a conservative effect size of 0.30. Considering a dropout rate of 15% and 5% of candidates who will not meet inclusion criteria, we will recruit at least 155 participants per

condition (N=310). The expected effect size and dropout rates come from studies showing that blended interventions lead to lower dropout rates [20] and better outcomes in patients with anxiety and adjustment disorder [18] when compared to face-to-face interventions."

Regarding suicidal risk and substance use, these will be assessed using the Mini-International Neuropsychiatric Interview (MINI). We have added this to the text.

Second, regarding the data collection, the use of the UP-APP to collect measures from the UP group guarantees a good ratio of completion within timing, provided a reliable app is developed; while I'm sure the authors have already considered this, a pilot study on the app is highly recommended. Also, it might have research interests by itself; we're currently lacking studies on the advantages of employing apps to collect data compared to traditional assessment. On the contrary, the TAU group might fare not so well; measures are planned to be collected weekly, but session frequency is left open to the "characteristics of their centres" (p. 11). Moreover, key information about the development of the TAU therapy is planned to be acquired at the end of the sessions. This poses some threat to the appearance of missing data and memory biases. I'm concerned that these issues might introduce problems in the comparison of both treatments, which might prevent solid conclusions being drawn on the role of the UP plus blended methodology due to disparities in the data available for both treatments, or the amount of therapy received. Also, I'm sure that the authors realize that the comparison between UP blended therapy and TAU makes it difficult to conclude the separate role of the UP protocol or the blended element in an eventual superiority of the UP plus app group.

Response: Thank you very much for your comments and suggestions. Regarding the pilot study, thank you for highlighting this idea because it has been a mistake do not mention it in the text. We have now added this to the text.

Pg. 9 Lines 4-16: "Previous to start the RCT we will conduct an open pilot study to analyze the preliminary data of the clinical utility and feasibility of the UP-APP in a small sample of patients with emotional disorders diagnosis. Specifically, after the clinical assessment, from those who met the inclusion and exclusion criteria, we will invite 10 patients (in order of date of receipt) to participate voluntarily in this pilot study. Participants will sign the informed consent and data protection. Then, they will be randomized to one baseline condition: 1, 2 and 3 weeks in order of date of receipt (baseline measures will be ODSIS and OASIS [22]). Then patients will receive a face-to-face psychological treatment in a blended format and will receive the instructions to download the UP-APP in their Smartphone. They will be asked to complete a special set of questions to assess the comprehension, appearance, utility, interest, if they would recommend it to other people, usability, intention to use in the future, and satisfaction of the contents of each module of the UP-APP (ad hoc)."

Regarding assessments:
On the one hand, the TAU group will not be able to be evaluated on a weekly basis. The weekly evaluation is only carried out in the UP group using the ODSIS and the OASIS. Both questionnaires, the ODSIS and the OASIS, are used as primary measures in the two conditions (in the pre-test, the post-test and the follow-ups) and also as a weekly assessment of treatment evolution in the UP group, as recommended by the original manual of the UP. On the other hand, while key information about the development of the TAU therapy is planned to be acquired at the end of the sessions, we have previously known that therapists in this condition will apply unstructured CBT interventions, using the following techniques: Psychoeducation, Cognitive restructuring, Relaxation techniques, Mindfulness techniques, Exposure techniques, Activity scheduling, Problem solving and training in Communication techniques. We have added this information in the text:

Pg. 10 Lines 22-25: “using the following techniques: Psychoeducation, cognitive restructuring, relaxation techniques, mindfulness techniques, exposure techniques, activity scheduling, problem solving and training in communication techniques.”.

Finally, we are aware of the difficulty explained by the reviewer. However, the aim of this naturalistic study is to find out whether what is offered as an alternative is effective in obtaining a statistically significant improvement over the TAU group. In future studies, it would of course be interesting to analyze the differential role of blended therapy, but this would require a less naturalistic and more controlled approach which is out of the scope of the present work.

Third, the last observation carried forward (LOCF) strategy to approach missing data mentioned in p. 12 has been criticized as lacking a sound statistical basis (Leon et al., 2006; Sanz & García-Vera, 2013). I suggest using only modelling techniques for missing data imputation whenever possible, given an acceptable percentage of missing data (less than 20-30%) and a random pattern of missing data.

Response: Thank you very much for the suggestion. We completely agree with the reviewer. In order to avoid performing the last observation carried forward (LOCF), we are going to carry out linear mixed model analysis since these analyses allow us to work with missing values (Krueger & Tian, 2004). We have removed the references to the LOCF from the article.

Pg 13 Lines 13-14: Missing data will be handled using mixed models, which can effectively handle missing data [51].

Finally, I consider that the calculation of a Reliable Change Index (RCI) and a Reliable Recovery Index (RRI), as proposed by Jacobson and Truax (1991), would add a powerful yet simple tool to assess – and communicate – the effectiveness of both interventions, beyond the already powerful Cohen’s d effect size. Even with the employment of the MINI as recovery criteria, I encourage the authors to consider these two procedures for their data analysis strategy.

Response: We totally agree with the reviewer. We have included the Reliable Change Index (RCI) and a Reliable Recovery Index (RRI) in the analyses that we will carry out.

Pg 13 Lines 8-12: “These analyses will be computed both for the primary and the secondary outcomes. The effect sizes will be computed and interpreted following the Cohen's proposal. Additionally, we will also calculate the Reliable Change Index (RCI) and the Reliable Recovery Index (RRI) to evaluate the effectiveness of both interventions, as proposed by Jacobson and Truax [50].”

Reviewer: 2
I appreciate the opportunity to review the protocol: “Implementation, efficacy, and cost-effectiveness of the Unified Protocol in a Blended Format for the Transdiagnostic Treatment of Emotional Disorders: study protocol for a multicentre, randomized, superiority controlled trial in the Spanish National Health System.” This protocol presents a multicenter, randomized, superiority, clinical trial evaluating the UP transdiagnostic treatment by comparing CBT (which they define as treatment as usual) with a blended format UP (delivered through both in-person and online formats). I believe that this study could make significant contributions to the literature and our understanding of how blending in-person clinical services with app-based treatment may support patients and providers while alleviating burden for mental health systems. However, several clarifications and revisions should be made to the protocol, delineated below. I appreciate that the researchers intend for the study to be as naturalistic as possible, in order to ensure accurate representation of services delivered in public settings, however, as the authors indicate, this adds many confounding variables that may obfuscate study results.

Additionally, there seem to be other more significant confounds that make it difficult to disentangle whether differences between study condition are attributable to differences in treatment (CBT vs UP) or to the blending of in-person and online services. Finally, blended treatment is proposed as a way of alleviating burden on care systems and saving time for clinicians, but in this study online interventions seem to be more of a supplement to treatment than a replacement for some in-person sessions. Additional comments follow:

Abstract
The final bullet point in strengths and limitations is syntactically incorrect.
Response: We apologized for this mistake. The new bullet point is now described as follows: “One limitation could be that some people may be resistant to participate in the blended condition because they perceive it as more impersonal and less effective”.
Introduction
Pg 5 Line 47: Word “collapsed” should be replaced with “overwhelmed” or a synonym.
Response: Thank you for this suggestion. We have replaced the term to overwhelmed.
Pg 6 Line 3: misplaced infinitive “it allows to simultaneously treat” or missing noun
Response: As the reviewer suggested, we have corrected the phrase as follows: “it allows the simultaneous treatment of people with different EDs and comorbid presentations with a single protocol”.
The introduction would benefit from more background information on blended treatments. It states that blended treatments are “dynamic and flexible because they allow using technology to motivate, monitor, give support, and treat patients,” but these are a broad array of benefits and outcomes that seem to be referenced together without citing research that clearly demonstrates these benefits. They should be parsed out more, since flexibility for clinicians is a different issue from monitoring and motivation. Please include information about other treatments that have been blended, and whether the blending usually includes supplemental support or the replacement of some in-person sessions with online modules.
Response: Thank you for this comment. As the reviewer suggested, we have included information on blended interventions in a more specific way.
Pg 3 Lines 23-30: “Research has shown that blended interventions are more effective than face-to-face treatments in the reduction of depression and anxiety symptoms [18]. For example, one study found that a blended smartphone treatment, which consisted of four face-to-face sessions and a smartphone app to be used between the sessions, can be as effective as a full behavioural activation treatment in the reduction of major depression. Moreover, comparable scores were also obtained between the two conditions for treatment credibility and working alliance, and therapist time was reduced by an average of 47% in the blended condition [19].”
Pg. 6 line 40: “save time to the clinicians” should be reworded, and “w” should be replaced with a full word.
Response: Thank you for highlighting these mistakes. We have corrected them. It was a transcription error.”.
Pg 3 Line 32: “given that they allow saving time to the clinicians”
Cost-efficiency is referenced as a study outcome in the introduction and title but not in the abstract.
Response: Thank you for this comment. We have added cost-efficiency as a study outcome in the abstract: “Cost-efficiency of the intervention, App usability, as well as opinion and confidence in the treatment will also be evaluated. Assessment points will include baseline and 3, 6 and 12 months after treatment onset.”

Pg 6. 50: “our goal will be tested” should be reworded for clarity (e.g. “our outcomes will be evaluated,” “our aims will be assessed,” “our study will be conducted”)
Response: Thank you so much for this comment. As the reviewer suggested, we have replaced “our goal will be tested” for “our outcomes will be evaluated”.
Study Protocol

I am struck again by the significant confound of providing two different treatment protocols, one structured and one unstructured, but suggesting that the superiority will be attributable to the blended format, or additional treatment, in one condition. Why not offer the UP to both groups so that any comparative benefits would more clearly be attributable to the additional online support? As the study is currently described, differences in condition could be due to the protocol (UP vs CBT), the presence or lack of structured protocol, the addition of an online element, or simply additional support.
Response: Thank you for this comment. We consider that the UP in a blended format condition will be superior because participants will have access to therapeutic content for a longer period of time. Regarding the UP, it has been seen that this intervention is more effective in a group format, but groups are not always possible in routine care in Spain due to the difficulties in grouping individuals with the exact same format and group format is not the preferred format by patients (Osma et al., 2019). Therefore, the aim of our study is to compare the treatment that is currently offered in public mental health units with the blended format, thus setting a relatively naturalistic study. Of course, in future studies it would be interesting to analyze the differential role of blended therapy.

It would also be important to understand how therapists will be trained in the UP and how fidelity to the UP will be monitored to make sure the interventions (UP vs CBT) are, in fact, different. How will the researchers be ensuring that UP therapists are not providing UP to clients who are randomized to TAU?
Response: We have added a paragraph in the text about how therapists will be trained in the UP. To ensure that UP therapists are not providing UP to clients who are randomized to the TAU, the therapists assigned to the TAU condition will provide a description of the contents of the sessions. We believe it would be more appropriate to record the therapy sessions so that an external observer could ensure fidelity, but unfortunately this is not possible in a naturalistic context.

Pg 9, Lines 24-33: “Therapists in the UP group received a training workshop on UP prior to the start of the intervention. This consisted of 2 or 3 group workshop sessions in which the therapists were instructed on the delivery of the different UP treatment modules. The duration of the course was between 10 and 20 hours, depending on the availability of the therapists at the centre. In addition to the workshop, all therapists received individual training during 12 therapy sessions. The individual training consisted of either online supervision before each session or participation as a co-therapist with an expert in the implementation of the UP intervention, who also evaluates the fidelity of the treatment. In both cases, the training was led by the lead author (blind note), who has been certified as a UP Trainer by the Unified Protocol Institute.”
Line 20 “standard intervention” should be clarified as to which treatment group the authors are referring to.
Response: Thank you for this comment. We have replaced “standard intervention” for “TAU”.
Sample Size
I appreciate the description of how a sample size was generated. However, a dropout rate of 15% in a public mental health setting seems optimistic, given that reported dropout rates are typically double that. The authors should explain how they determined this dropout rate (is it specific to the centers where the study will be conducted, or to the Spanish healthcare system?).
Fernández D, Vigo D, Sampson NA, Hwang I, Aguilar-Gaxiola S, Al-Hamzawi AO, Alonso J, Andrade LH, Bromet EJ, de Girolamo G, de Jonge P, Florescu S, Gureje O, Hinkov H, Hu C, Karam EG, Karam G, Kawakami N, Kiejna A, Kovess-Masfety V, Medina-Mora ME, Navarro-Mateu F, Ojagbemi A, O'Neill S, Piazza M, Posada-Villa J, Rapsey C, Williams DR, Xavier M, Ziv Y, Kessler RC, Haro JM. Patterns of care and dropout rates from outpatient mental healthcare in low-, middle- and high-income countries from the World Health Organization's World Mental Health Survey Initiative. Psychol Med. 2020 Apr 28:1-13. doi: 10.1017/S0033291720000884. Epub ahead of print. PMID: 32343221; PMCID: PMC8265313.
Response: We appreciate the reviewer's comment. As we explained in a similar question from reviewer 1, in the sample size calculation we have taken into account previous research showing that

dropout rates are significantly lower in blended interventions (Erbe et al., 2017). Furthermore, we considered that overestimating the sample size required for a study might also be detrimental in terms of clinician/researcher/participant burden, which is why we have relied on previous literature to calculate the sample size and provide an achievable, yet arguably sufficient sample size to achieve our goals. We now describe this in more detail, with adequate references, in the text.

Procedure

How will the qualitative analysis of focus group data be collected? Under sample and recruitment, which mental health professionals will be responsible for assessing diagnoses? The referring ones or study therapists?

Response: Thank you for these questions. In response to the first, we have added more information in the text about the focus group procedure.

Pg 6 Lines 11-15: “The focus groups will be recorded on video to be transcribed by two researchers of the study. The qualitative analysis of the data collected will be used to design the UP-APP for Smartphone. This analysis will consist of generating a system of codes, grouping the he information provided by the participants in the focus groups that referred to the same ideas or highlighting the main ideas.”

In response to the second question, therapists and psychiatrists from the units to which patients are referred to and who want to collaborate in the study will be responsible for the assessment of diagnoses. We have added more information in the manuscript to clarify this point.

Pg 7 Lines 7-10: Mental health professionals (therapists and psychiatrists from the units to which patients are referred to and who want to collaborate in the study) will be responsible for assessing the current DSM diagnoses (See “Measures” section) and the remaining eligibility criteria (see “Eligibility criteria” section).

Under randomization, the section about patients who refuse to participate in the study is confusing. Does this mean you will be collecting data on these patients as well?

Response: Thank you for this comment. We agree that the sentence could be confusing in the way it is described in the manuscript. We mean that we will collect the data about how many people refuse to participate and why. This information can be of interest for future studies, as we indicated in the text. We have rewritten the sentence as follows:

Pg 8 Line 5-7: “Patients who refuse to participate in the study will receive the TAU outside the RCT. The number of people refusing to participate and the reasons for that decision will be recorded and reported due its interest for future studies”

Regarding the stratified sampling strategy, how will the researchers account for individuals who fit into multiple diagnostic categories (i.e. GAD and depression) and are more severe in one or less severe in another? Please be more specific about “the recommended cut-off in the manuals.” Which measures are you using to determine severity of depression and/or anxiety, and what cut-offs will determine severity for stratified randomizing? How will researchers be simultaneously ensuring that 10 participants at each cite are assigned to each group (as is indicated in the flow chart).

Response: While considering that the comorbidity of the patients could be a good strategy to conduct the stratification, we have preferred to consider the severity of the primary symptoms, thus anxiety (OASIS) and depression (ODSIS). This strategy has been successful in a previous RCT implemented by our team in the Spanish mental health system (Osma et al., 2021). The cut-off reported in Spanish clinical samples of persons with EDs has been 10 (0-20 range) in both scales (Osma et al., 2019). This cut-off differentiates patients with moderate-severe symptoms from those with moderate-low symptoms. We will randomize the same number of patients above and beyond this score for the ODSIS and the OASIS. We have added more information in the text in this regard.

Pg 8 Line 10-13: “cut-off reported in Spanish clinical samples of persons with EDs, which has been 10 (0-20 range) in both scales [22]. This cut-off differentiates patients with moderate-severe symptoms from those with moderate-low symptoms.”

Thanks so much for underlining this mistake in the Flow Chart regarding the number of patients assigned to each group. It’ll be easier to complete the participants in each condition as soon as they

met the inclusion criteria and have been randomized to one of the two conditions. We have removed the incorrect information in the flow chart.
Pg. 11, line 16: typo: “this individual face-to-face appointments” (these).
Response: Thank you for highlighting this mistake. As the reviewer suggested, we have replaced “this” for “these”.
Pg. 11 line 19-20: typo in the following sentence: “clinicians will recommend the participants in the blended condition to work on…”
Response: Thank you for this comment. We have corrected the typo in the sentence.
Pg 8 Lines 26-29: “Clinicians will recommend participants in the blended condition to work on modules 1, 2, 5, 6 and 8 during at least one week, and modules 3, 4 and 7 during at least two weeks (see the “Unified Protocol in a blended format” section for a detail on the titles of the UP modules)”.
p. 11 line 26: should read “prevents us from defining”
Response: Thank you for this comment. We have reworded the phrase as the reviewer suggested.
In your description of the two treatments, pharmacological treatment is mentioned only under TAU. The inclusion criteria chart suggested that nobody would be excluded due to psychotropic treatment, but it is confusing that this is mentioned only under TAU. Presumably participants in either group may be receiving conjunctive pharmacological treatment. Please clarify.
Response: Thank you again for this comment. We agree that this information should be clarified in the text. We have separated this information from the TAU description. The text now reads as follows:
Pg 8 Line 19-21: “Individuals with an ED also frequently receive pharmacological treatment (i.e., antidepressants and / or anxiolytics) as the treatment of choice in the Spanish Mental Health System.
UP treatment group: The team has obviously not developed the app yet, and plans to design it in accordance with expressed preferences gathered in focus groups, but more information about what the app might involve would be useful. Will it have daily mood tracking or activities related to each UP module? Will it have UP homework assignments in digital format? For readers unfamiliar with the UP module, please provide examples of the types of engagement users will have with the app and more information about what the UP entails.
Response: Thank you for this comment. We have added more information about what the app will involve in the text.
Pg 9 Lines 9-20: “In the UP-APP, after completing each module, an assessment of the knowledge acquired will be carried out using true/false questions. The App will collect the correct/incorrect responses and will provide feedback to the participants. Thus, participants will receive positive reinforcement as they progress through the modules and get correct answers to keep them engaged and motivated in the use of the App. In addition, participants will have to complete different exercises throughout the modules, such as records or activities to identify emotion-driven behaviours. They will also be provided with examples of real patients with whom they can identify and which will help them to complete their records. Finally, a weekly assessment will be made to evaluate the evolution of the depression and the anxiety symptoms (ODSIS and OASIS) [22]. The scores over time will be shown to the participants with a graphic display. This weekly evaluation with the APP will also include the participants´ degree of motivation to continue working on the intervention.”

Measures
More information about all chosen measures would be beneficial to this protocol, including information on their validation, reliability, and general descriptions of the measure. For example, are the SUS and CEQ self-reports? Are they completed by therapists or administrators? What is the CSRI? Given the large number of outcomes being evaluated, the measures section would benefit from being subdivided into clearer sections (i.e. secondary outcomes could include Patient Outcomes, Implementation Outcomes, and Patient Satisfaction Outcomes).
Response: Thank you for this suggestion. As recommended by the reviewer, we have subdivided the measures section and added more information about all the chosen measures. In order to optimize space in the text we have placed the information in a new Table (Pg 12 and 13).
App Outcomes

I believe that “without the need to ask the participant” intends to clarify that participants will not need to actively respond to questions about their use, but it reads as a lack of consent. I assume that participants will be informed of all data gathered by the app. Please clarify.
Response: Yes, this is what we meant. We have now clarified this information by adding the following sentence:
Pg 11 Line 12-14: “All the participants using the UP-App will be informed about the data that is going to be registered while using it.”
Analyses
Given the large number of possible confounds in this study (i.e. number of sessions, length of treatment sessions, frequency of appointments) and the large number of outcomes collected, what covariates will be included in your analyses? In the multi-level model, what will each level of the model be? Will models account for treatment center where interventions were received?
Response: We appreciate the reviewer's interest in detailing how the different variables in this study will be analyzed. The linear mixed model analysis will include the main effects of time (each variable collected at each evaluation time to analyze the evolution over time). The treatment condition and the number of sessions will also be included as interaction effects with time (in order to see differences in the evolution of the variables as a function of the treatment condition and/or as a function of the number of sessions). Finally, as suggested by the reviewer, the linear mixed model analysis will also include the variable of the center where the participants have received the treatment as random effects in the model. We have added this information in the manuscript:
Pg 14 Lines 7-13: “Specifically, the linear mixed model analysis will include the main effects of time (each variable collected at each evaluation time to analyze the evolution over time). The treatment condition and the number of sessions will also be included as interaction effects with time (in order to see differences in the evolution of the variables as a function of the treatment condition and/or as a function of the number of sessions). Finally, the center where the participants have received the treatment will be included as random effects in the model.”
Please provide further information on how relationship between cost of intervention and QALYs will be analyzed.
Response: Thank you very much for the suggestion, we have added more information for the relationship between cost of intervention and QALYs.
Pg 13 Lines 15-23: “Costs-effectiveness will be calculated by exploring the relationship between the cost of each intervention (cost of TAU or UP in a blended format, number of sessions, medication and use of health resources carried out by the participants [evaluated through the CSRI]) and its consequences in the form of QALYs (standardized health units that allow the quantification of individuals' preferences regarding the quality of life that has been produced by a health intervention [52]. The information obtained with the Euroqol allows the calculation of QALYs). Other measures of intervention penetration will be used, such as the number of consumers who were eligible or willing to use the app (end users).”
Conclusions
The study is presented in the conclusions as offering insight into a potential solution to long waiting lists, however this study does not include treatment time or other measures related to waitlists as primary study outcomes. This seems like a future direction rather than something that the study will “reveal.”
Response: Thank you very much for this comment. As the Editor suggested, we have removed the conclusions section.
Appendices
Pg. 23: Flowchart: This chart should have a title and heading.
Response: We have corrected this as indicated.
The table of assessments included an “intervention” assessment (t2) where the ODSIS and OASIS are administered. When will this occur? Or will these be provided throughout treatment, and if so how frequently? The study protocol suggests that these two measures will be delivered to APP users through the App. How will the TAU group receive these measures throughout treatment?

Response: Thank you for this comment. The ODSIS and the OASIS will be administered as primary measures in both conditions at pre, post and follow-ups. These measures will also be administered through the App on a weekly basis only in the UP condition as recommended by the original UP manual.

References

Erbe, D., Eichert, H.-C., Riper, H., & Ebert, D. D. (2017). Blending Face-to-Face and Internet-Based Interventions for the Treatment of Mental Disorders in Adults: Systematic Review. Journal of Medical Internet Research, 19(9). https://doi.org/10.2196/jmir.6588

Krueger, C., & Tian, L. (2004). A Comparison of the General Linear Mixed Model and Repeated Measures ANOVA Using a Dataset with Multiple Missing Data Points. Biological Research For Nursing, 6(2). https://doi.org/10.1177/1099800404267682

Leterme, A. C., Behal, H., Demarty, A. L., Barasino, O., Rougegrez, L., Labreuche, J., Duhamel, A., Vaiva, G., & Servant, D. (2020). A blended cognitive behavioral intervention for patients with adjustment disorder with anxiety: A randomized controlled trial. Internet Interventions, 21. https://doi.org/10.1016/j.invent.2020.100329

VERSION 2 – REVIEW

REVIEWER	Estupiñá-Puig, Francisco Complutense University of Madrid, Dpt. of Personality, Evaluation & Clinical Psychology
REVIEW RETURNED	30-Oct-2021

GENERAL COMMENTS	I'd like to thank the authors for considering my suggestions and clarifying my doubts. I think the manuscript has improved and increased clarity is shed upon the details of sample size calculation, assessment of exclusion criteria and statistical analysis. I also consider that the details of the app for the blended condition are clearer now and the inclusion of the pilot study seems just right. Also, I understand that the authors intend to exercise caution when comparing both interventions, limiting themselves to their effectiveness and efficiency. I hope that in future studies the team will consider developing dismantling studies to analyze the precise role of the app and the structured nature of the intervention on both effectiveness and efficiency.

VERSION 2 – AUTHOR RESPONSE

Reviewer: 1

Dr. Francisco Estupiñá-Puig, Complutense University of Madrid

Comments to the Author:

I'd like to thank the authors for considering my suggestions and clarifying my doubts. I think the manuscript has improved and increased clarity is shed upon the details of sample size calculation, assessment of exclusion criteria and statistical analysis. I also consider that the details of the app for the blended condition are clearer now and the inclusion of the pilot study seems just right. Also, I

understand that the authors intend to exercise caution when comparing both interventions, limiting themselves to their effectiveness and efficiency. I hope that in future studies the team will consider developing dismantling studies to analyze the precise role of the app and the structured nature of the intervention on both effectiveness and efficiency.

Reviewer: 1

Competing interests of Reviewer: I declare no competing interests

Response: Thank you so much for your kind words. We really appreciate your comprehension regarding the comparisons in our study and we keep in mind for future studies to conduct dismantling studies which will also answer interesting questions regarding effectiveness and efficiency of the app.